The Sun™ StorEdge™ A5000 Architecture

Technical White Paper

© 1997, 1998 Sun Microsystems, Inc. All rights reserved.

Printed in the United States of America.
901 San Antonio Road, Palo Alto, California 94303 U.S.A

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

TRADEMARKS
Sun, Sun Microsystems, the Sun logo, Solaris, NFS, SunVTS, Sun Enterprise Volume Manager, StorEdge, Solstice, and Solstice SyMON are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries.UNIX is a registered trademark in the United States and other countries, exclusively licensed through X/Open Company, Ltd.

All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the United States and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

THIS PUBLICATION IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS PUBLICATION COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THE PUBLICATION. SUN MICROSYSTEMS, INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS PUBLICATION AT ANY TIME.

Please
Recycle

Adobe PostScript

Contents

1. The Intelligent Storage Network Revolution 1

Sun's Intelligent Storage Network Vision 1

The Sun StorEdge A5000 3

Meeting the Needs of Today's Businesses 5

Scalable Storage from Sun 5

2. Sun StorEdge A5000 Enabling Technologies 7

Fibre Channel. 7

RAID 13

Dynamic Multi-Pathing 16

3. The Sun StorEdge A5000 Architecture 17

Capacity 17

System Overview 18

Fibre Channel Interface 19

Gigabit Interface Converter (GBIC) Module. 20

Interface Board. 21

Drive Technology. 22

Host Adapter 23

Enclosure Services and the Front Panel Module 25
Enclosure and Power 26
Fan Trays 28
FC-AL Hubs 28
Designed for Reliability, Availability, and Serviceability 29
Performance 32
Scalability 33
Specifications and Standards 34
4. Sun StorEdge A5000 Software 35
Sun Enterprise Volume Manager™ 35
luxadm 38
Solstice™ SyMON™ 39
SunVTS™ 39
5. Application Environments 41
On-line Transaction and Database Processing 41
Data Warehousing and Decision Support 42
Network File Service 43
Imaging 43
Interactive Services and Multimedia 44
Clusters 44
6. Summary 47
A. Glossary 49
References 55

The Intelligent Storage Network Revolution 1

Today's businesses are information driven. The need to access and analyze corporate information in real-time, update databases, perform trend analysis, provide high customer satisfaction, and operate in 7x24 environments is changing the demands placed on storage systems. It is no longer sufficient for mass storage subsystems to simply provide increasing levels of capacity — they must also be fast, available, reliable, and highly serviceable in order to meet the requirements of both users and applications.

Traditionally, storage systems were designed as an adjunct to the computing environment, with new protocols like the Small Computer Systems Interface (SCSI) being created and modified as performance needs dictated. Perhaps the most popular add-on peripheral protocol to date, SCSI is reaching its performance and architectural limits. Combined with their other liabilities, distributed SCSI storage systems are becoming a severe bottleneck as computer systems and networking technology continue to advance and the demand for fast data access grows.

Sun's Intelligent Storage Network Vision

These trends underscore the need not only for recentralizing shared data, but also for ensuring that data can be accessed by a wide variety of users quickly and continually. By combining the latest storage technology breakthroughs with high speed networking, organizations can create a scalable Intelligent Storage Network — an environment in which information and administration is centralized. Dedicated to storage, an Intelligent Storage Network offers many of the features associated with today's networks within a framework

designed to meet changing data storage requirements. Like the best networks, an Intelligent Storage Network is standards-based, scalable, modular, multi-pathed, centrally managed, multi-vendor, and highly available. Like the best peripherals, it is also high performance, highly available, low latency, configurable for a variety of workloads, and able to perform well in clustered environments (Figure 1-1).

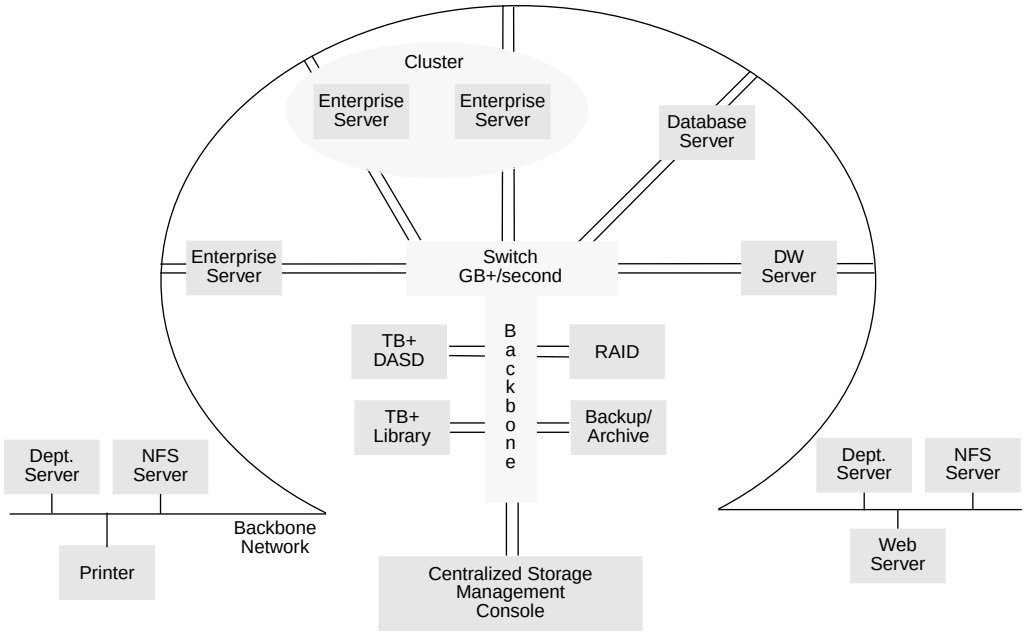

Figure 1-1 Sun's Intelligent Storage Network vision

Sun Microsystems, Incorporated believes that Fibre Channel is the core technology that enables the Intelligent Storage Network to become a reality. The flexibility, high performance, and reliability of Fibre Channel technology enables organizations to create big, fast storage networks into which not only disks and RAID subsystems can be plugged, but eventually tape backup, archive, hierarchical storage management (HSM), and library systems as well. No longer will the enterprise be constrained to move data at T1, T3, ESCON, SCSI, and Fast Ethernet speeds — significantly faster pipelines based on Fibre Channel technology will become available.

Just as networks grow more valuable as more people and applications are connected to them, storage networks grow more valuable as their capacity increases. As the Intelligent Storage Network grows, it will feed large server systems, which in turn will provide slower corporate backbone networks with the information that knowledge workers and local applications need. Using this technology, organizations can process, organize, and create the structured information needed to ensure the enterprise remains competitive.

The Sun StorEdge A5000

The Sun™ StorEdge™ A5000 is the first intelligent storage network product in a new product family from Sun. Using second generation Fibre Channel technology and offering high reliability, availability, and serviceability (RAS) features, the Sun StorEdge A5000 is scalable from the desktop to the datacenter — offering higher performance than any other product in its class (Figure 1-2).

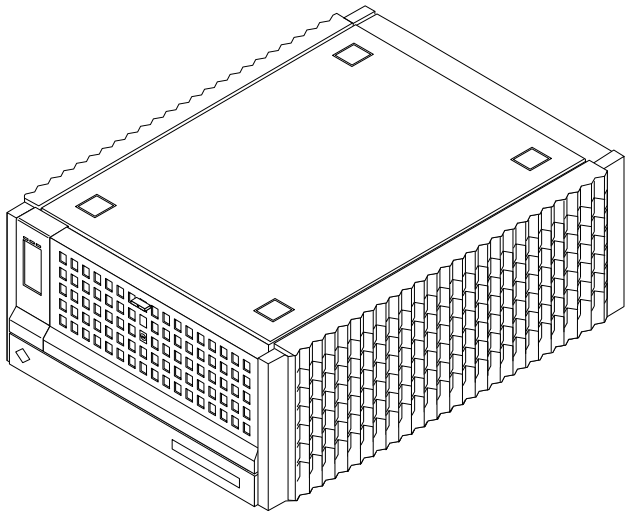

Figure 1-2 The Sun StorEdge A5000

1

One of the most important new storage products from Sun in years, the Sun StorEdge A5000 offers the capacity, performance, and reliability, availability, and serviceability features users demand in an economical package (Table 1-1).

	Features
Capacity	• 45 GB to over 12 TB configurations
Disk Drives	• Half-height, 9 GB, 3.5” Fibre Channel Arbitrated Loop disk drives • Dual-ported drives • Fibre Channel Arbitrated Loop (FC-AL)
RAID	• RAID levels 0, 1, 0+1, 5 • Host-based RAID
Performance	• 100 MB/second full-duplex Fibre Channel technology • Over 10,000 I/O operations per second (IOPS) per loop • Over 95 MB/second user data bandwidth per loop • Load balancing across loops
RAS	• Dynamic Multi-Pathing support • Automatic loop failover • Dual ported FC-AL disk drives • Hot relocation support • Redundant, hot-swappable power • Redundant drive interfaces • Hot-swappable drives • Redundant, hot-swappable array interfaces • Hot-swappable cooling units • Full spectrum of on-board and host-based diagnostics • Full CRC datapath support • Front panel module and electro-luminescent touch screen and display
Software	• Solaris™ 2.5.1 (8/97 or later) or Solaris 2.6 • Sun Enterprise Volume Manager 2.5 • Command line and scripting utility (luxadm) • Integration with Solstice™ SyMON™
Configurations	• Tabletop and cabinet configurations

Table 1-1 Sun StorEdge A5000 features

Meeting the Needs of Today's Businesses

Designed for capacity, performance, reliability, availability, and serviceability, the Sun StorEdge A5000 meets the mass storage needs of users in a numbers of disciplines:

- On-line transaction processing (OLTP)
- Data warehousing
- Decision support
- Seismic analysis
- Video production
- Mechanical design (MCAD)
- Web service
- Network file service
- Imaging

Able to deliver high performance and high RAS at low cost, the Sun StorEdge A5000 is a scalable architecture that is ideal for cost-sensitive or volume applications where fast data access is required. With both tabletop and cabinet configurations, the Sun StorEdge A5000 is a flexible mass storage subsystem that supports departmental, datacenter, and clustering environments with ease.

Scalable Storage from Sun

Today, competitive pressures and demands for financial accountability have forced organizations to seek information systems architectures that can be easily deployed and relocated as business conditions warrant. There is no doubt that flexibility and investment protecting scalability are essential requirements for modern computer equipment. For over a decade, Sun has responded to these demands with a line of open, networked, binary-compatible workstations and servers based on SPARC™ Reduced Instruction Set (RISC) processors.

Storage systems are a key element in providing balanced performance, and must be tuned to complement the performance and scalability of powerful multiprocessor systems like those offered from Sun. Utilizing second generation Fibre Channel technology, the Sun StorEdge A5000 is the first in a series of breakthrough product offerings in high performance, high RAS storage systems that scale from workgroups, to departmental, to datacenter, to mainframe class environments.

1

With carefully balanced performance, a full complement of advanced hardware and software features, and full compatibility with existing Sun desktop, server, and storage systems, the Sun StorEdge A5000 is a potent expression of next-generation technology that promises to usher in a new era in intelligent network storage.

Sun StorEdge A5000 Enabling Technologies

The Sun StorEdge A5000 uses a combination of Fibre Channel, RAID, and Dynamic Multi-Pathing technologies to create a powerful, flexible, and scalable mass storage architecture that provides unprecedented levels of capacity, performance, reliability, availability, and serviceability.

Fibre Channel

Sun strongly believes in the power, flexibility, and economy of independent servers connected to clients through powerful networks. As the industry embraces this model of computing, Sun is turning its attention to a new generation of intelligent storage networks that are more autonomous and independent. Sun also believes that conventional peripheral interface technologies are inadequate for future systems. SCSI, for instance, has many shortcomings, including limited bandwidth, restrictive distances between devices, expensive and complex cabling, and impedance matching problems.

Fibre Channel is a high performance interconnect standard designed for bidirectional, point-to-point communications between high performance workstations, peripherals, and large host systems. It offers a variety of benefits over other link-level protocols:

- Efficiency and high performance
- Scalability
- Operation over long distances
- Easy to cable
- Supports popular high level protocols
- Supports switched networks

Theory of Operation

Channels are a generic class of communications technology designed to move data from one point to another with high speed and minimum latency. For speed, functions such as error correction and retry-on-busy are typically implemented in hardware. Because they do not have sophisticated software-based routing and switching capabilities, channels typically operate in environments where every device on the channel is defined and known in advance.

Networks, unlike channels, have tremendous flexibility in managing communications with other devices on the network. Software intensive, networks can make complex routing decisions to overcome broken connections, noise, or other difficulties. Because of this dependence on high-overhead, software-based decision making, network performance typically falls far below the theoretical maximum performance of the network media.

At its most basic, Fibre Channel is just a high speed serial connection between two buffers. Indifferent to the format of the data, Fibre Channel only provides the error checking and necessary signaling to ensure the reliable movement of data between the source and destination buffers.

Fibre Channel employs a mechanism known as a *fabric* to establish connections between *ports*. The fabric, like the telephone network, has all the needed intelligence to make routing decisions. The only concern of a port is to ensure its proper connection to the fabric. Each fabric has an address space that allows up to 16 million addresses. A fabric can be simple, or a mix of switches and hubs that connect a complex and widely dispersed network of ports.

Reliability and Throughput

There are no limits on the size of a transfer between applications using Fibre Channel. Each sequence is broken into frames, the framesize being negotiated by the ports and the fabric. A 32-bit cyclic redundancy check ensures that transmission errors can be reliably detected.

Unlike SCSI, IPI, FDDI, and similar transport level protocols, Fibre Channel is defined to be full-duplex, allowing data transfers at channel speeds in both directions simultaneously. In a balanced application, a 100 MB/second Fibre Channel connection can potentially move 200 MB/second of data.

Fibre Channel Arbitrated Loop

An important enhancement to Fibre Channel has been the development of Fibre Channel Arbitrated Loop (FC-AL), developed specifically to meet the needs of storage interconnects. Employing a simple loop topology, FC-AL can support both simple configurations and sophisticated arrangements of hubs, switches, servers, and storage systems (Figure 2-1). Furthermore, by using SCSI protocols over the much faster, more robust Fibre Channel link, FC-AL provides higher levels of performance without requiring expensive and complex changes to existing device drivers.

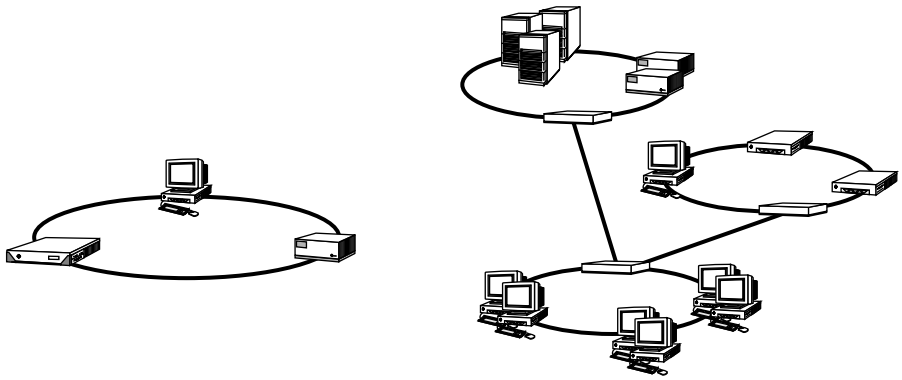

Figure 2-1 FC-AL's loop topology can support both simple and complex configurations.

Key Benefits of FC-AL

FC-AL offers a host of benefits over other protocols, enabling storage systems based on it to foster the creation of new applications that take full advantage of some impressive specifications:

- *Gigabit bandwidth*

  FC-AL is capable of data transfer rates of up to 100 MB/second (dual ported), with 400 MB/second envisioned for the future — far faster than SCSI, Serial Storage Architecture (SSA), or P1394 (Firewire).

- *Suitability for networks*

  In addition to performance, Fibre Channel is ideal for building networks. Employing hubs and switches similar to those used in networks, Fibre Channel enables complex arrangements of storage and systems to be connected together in highly scalable, highly available networks, or fabrics.

- *Use of existing SCSI protocols*

  FC-AL allows SCSI command packets to be sent over a high speed physical medium, reducing software and firmware costs and minimizing impact on existing software.

- *Node addressability far better than SCSI*

  With the ability to support up to 126 FC-AL devices on a single host adaptor, cost and implementation complexity is greatly reduced. Using optical fiber media, a single FC-AL loop can support nodes with a separation of up to ten kilometers from the host.

- *Unique device naming convention*

  FC-AL employs unique, 64-bit world wide names for devices. Disk drives, host adapters, and interface boards are identified by a world wide name, enabling improved diagnostics and software support.

- *Greatly simplified wiring and cabling requirements*

  Because Fibre Channel is a largely optical, simple serial protocol, electrical interference and expensive cabling are much less of an issue than with the complex parallel data paths used by SCSI.

In addition, FC-AL implementations support redundant data paths, hot pluggable components, multiple host connections, and dual ported drives — features that 15 year-old SCSI technology was never intended to support.

The technical advantages of FC-AL alone would be enough to convince most that it clearly represents the future of high speed peripheral interconnects, but FC-AL can also provide peace of mind to those who worry about the bottom line:

- *Industry standard*

  The FC-AL development effort is part of the ANSI/ISO accredited SCSI-3 standard, helping to avoid the creation of non-conformant, incompatible implementations.

- *Broadly supported*

  All major system vendors (Sun, HP, DEC, Compaq, and others) are implementing FC-AL, as are all major disk drive and storage subsystem vendors. The Fibre Channel Association, an industry group dedicated to the promotion of Fibre Channel, is a *Who's Who* of systems, subsystems, drive, and component vendors. Such wide support ensures competition, lower costs, and user choice.

- *Vastly more flexible*

  Fibre Channel can also be used to do more than disk I/O — the Fibre Channel specification supports high speed system and network interconnects using a wide variety of popular protocols, including HIPPI, TCP/IP, IPI, FDDI, ATM, as well as SCSI. Nearly all of the interconnect requirements of even the largest enterprises will one day be met by Fibre Channel, promising lower costs, easier administration, and the easy deployment and redeployment of computing resources.

Configuration Flexibility

Fibre Channel Arbitrated Loop enables storage systems to implement several loop configurations, each designed to address varying levels of performance and availability. Utilizing bypass circuits and multiplexors, FC-AL based storage devices like the Sun StorEdge A5000 permit the construction of multiple loops within a single enclosure, improving both performance and availability.

Environments needing the highest degrees of availability can employ a single loop for primary data access, enabling the second series of ports to function as a standby loop thereby increasing availability in the event of a loop failure. Systems that demand the highest performance, however, may choose to implement dual loops, mitigating the potential bottlenecks associated with a large number of drives sharing a single loop structure.

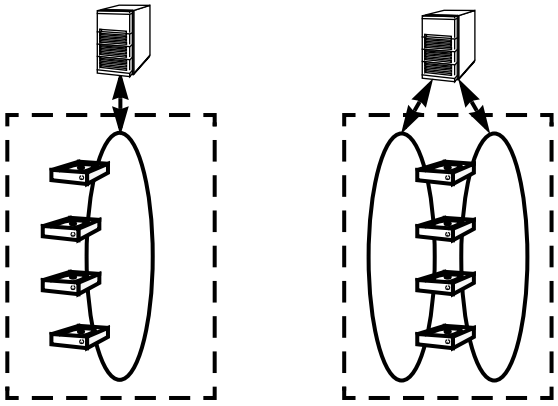

Figure 2-2 FC-AL's loop topology supports both single and dual loop configurations.

Fibre Channel Protocol

There are two aspects to establishing communication between a host and a peripheral — the medium which connects them, and the protocol with which they communicate. The Sun StorEdge A5000 employs Fibre Channel as its delivery medium, and a SCSI-3 protocol known as SCSI-FCP with which commands and data are transferred to the device.

Although the Sun StorEdge A5000 has a Fibre Channel interface, its programming model is that of a SCSI device. It supports SCSI through a protocol known as SCSI-FCP. Using this approach, existing SCSI drivers can send data in a serial stream compatible with lower level Fibre Channel framing and signaling protocols.

A library of commands is available to enable the host system to interrogate and configure the array across the Fibre Channel interface. Special commands enable control over internal logging and diagnostics. Individual drives can be accessed through an addressing scheme that specifies a controller number, a loop number, and target drive ID. In addition, there are special interfaces between the array interface board and the Sun Enterprise Volume Manager software to ensure optimal performance and minimal system overhead.

RAID

Redundant Arrays of Inexpensive Disks, or RAID systems, were first formally defined by researchers at Berkeley in 1987. The most important part of their idea was to combine the power of small, inexpensive drives to achieve performance equal to a single large, expensive disk. Because RAID required multiple spindles, features were also needed that could protect against the data loss that could be caused by the failure of any one of the drives. These original safeguards subsequently became an important feature in their own right, with many users purchasing RAID systems just for their availability.

Today, disk arrays are sold as an intelligently managed collection of disk drives organized to optimize performance for a particular set of tasks. Some of the benefits delivered by disk arrays employing RAID technology include:

- Higher transfer rates
- Large number of I/O transactions per second
- Greater on-line storage capacity
- Increased data availability and system reliability
- Easier management of large amounts of data
- Reduced maintenance and downtime

All RAID systems achieve higher performance and/or reliability by replicating or spanning data across multiple disks. Exactly how this is done has profound effects on subsystem performance, cost, reliability, and data recovery. Five RAID levels, numbered 1 through 5, were defined by the Berkeley researchers. Since that time, a few more have been added — some simply variations of the original five. The RAID levels offered in the Sun StorEdge A5000, levels 0, 1, 0+1, and 5, provide varying degrees of availability with corresponding trade-offs in performance and cost.

Level 0 — Striping

RAID-0, or striping, interleaves data across multiple disks to achieve high performance. Striping is an industry term for breaking a data stream up and placing it across multiple disks in equal-sized *chunks* or *stripe blocks*. Each chunk is sequentially written to successive drives in the array. The set of sequential chunks that begins with the first drive and ends with the last drive forms the “stripe”. Array management software running on the host, on a

special controller inside the array, or some combination of both, is responsible for making the array look like a normal, or *virtual*, disk to the operating system and applications.

Traditionally, each disk in a stripe is generally assumed to be on its own independent data channel, allowing the transfer rate of a RAID-0 implementation to approach the sum of the transfer rates of each of the drives. (Presuming that the intervening software and hardware has sufficient throughput.) Although striping is typically utilized to help achieve high performance, the use of 100 MB/second FC-AL on subsystems like the Sun StorEdge A5000 minimizes the need for multiple channels.

RAID-0 can also provide excellent I/O operations per second (IOPS) performance, because data is spread across many independent spindles and actuators, helping to balance the I/O load.

Level 1 — Mirroring

Used for a long time in fault tolerant applications, this technique is also called *shadowing* or *mirroring*. Designed to optimize data availability rather than speed, RAID-1 duplicates each write transaction to one or more “mirror” disks.

Mirrored systems can be used to improve the IOPS performance of read operations, because the least busy drive can be selected to service requests. Because all drives must be involved in write operations, write performance may be slightly worse than an independent drive.

RAID 0+1 — Striping plus Mirroring

Useful configurations have been created by combining two existing RAID levels. One of the more popular is known as RAID 0+1, which combines the reliability of mirroring (RAID 1) with the performance of striping (RAID 0).

The reliability of a striped and mirrored system is excellent because of the high redundancy afforded by mirroring. RAID 0+1 systems can tolerate the failure of one or more disks, and continue to deliver data with virtually no performance degradation, unlike RAID 5. RAID 0+1 systems do carry, however, the higher cost of mirrored systems, as data requiring protection needs more disk space than simple independent spindles.

Level 5 — Striping with Distributed Parity

RAID-5 exploits the performance gains of striping while mitigating its vulnerability to data loss by adding error correction information (called *parity*) to the data.

RAID-5 supports independent access to individual drives in a group with striped data. If the data is interleaved across the group of drives in blocks equal in size or larger than most I/O requests, then drives can respond independently to requests, and even to multiple requests simultaneously. This has the effect of allowing RAID-5 to have very good random read performance. Conversely, if the block sizes in a stripe are smaller than the average size of an I/O request, forcing multiple drives to work together, RAID-5 systems can also enjoy good sequential performance.

RAID-5 employs parity information to secure data in the event of a disk failure. With RAID-5, however, both the parity information and data are interleaved across the array in a cyclic pattern. If one of the disks fails, a simple algorithm enables the missing data to be recovered from the remaining disks. RAID-5 is a good compromise for those needing good random performance without incurring the expense of mirroring.

If RAID-5 has a failing, it is its poor performance on small writes. When a single-block write command is issued to a RAID-5 configuration, the system must first complete several I/O operations and parity calculations in a *read-modify-write* sequence, degrading random write performance.

Host-Based RAID versus Controller-Based RAID

In recent years, much has been said about the implementation options available for RAID subsystems, and contention has arisen over the perceived benefits and drawbacks to both host-based RAID and controller-based RAID devices. While traditional controller-based RAID subsystems employing SCSI technology were long seen as the answer to higher performance by minimizing the amount of traffic flowing between the host and peripheral, the advent of 100 MB/second Fibre Channel is bridging the gap. Utilizing a faster transport medium like Fibre Channel, the performance of host-based RAID storage systems compete with, and in many cases surpass, their controller-based RAID counterparts. In addition, the lower cost and higher degree of flexibility offered with host-based RAID solutions like the Sun StorEdge A5000 ensures that users have access to the best price/performance configurations available.

2

Dynamic Multi-Pathing

The use of FC-AL, dual-ported drives, and multiple loop configurations offers unprecedented levels of configuration flexibility and system availability. With this added flexibility comes the complexity and increased probability of a connection failure. No longer are a host and peripheral connected by a single, simple cable. Today, hubs, switches, interface converters, cables, and drives are interconnected in a network-like structure, and sophisticated software is needed to take advantage of its benefits as well as ensure proper behavior during both nominal and exception conditions.

In systems like the Sun StorEdge A5000, the ability to connect a single dual-ported drive to two FC-AL loops fosters a high degree of data availability. By supporting multiple host connections, it is possible for a given host to communicate with an FC-AL drive through one or more device paths. This has the advantage of enabling data transfers to be rerouted through an alternate should one device path fail, ensuring data availability.

It is critical, however, to ensure the host utilizes only one device path at a given time. Dynamic Multi-Pathing (DMP) accomplishes this task, eliminating the possibility of accidental data corruption by coordinating access to device paths, as well as enabling load balancing and failover protection. The Dynamic Multi-Pathing scheme also provides device path failover, and is completely transparent to the user and application, allowing existing environments to take advantage of its increased availability features without modification.

The Sun StorEdge A5000 Architecture

The Sun StorEdge A5000 was designed to provide high performance, reliability, availability, serviceability, scalability, and flexibility at very low cost. The use of high volume storage components, a flexible enclosure design, and standard Fibre Channel technology has resulted in greatly improved serviceability, high reliability, and low cost without compromising access to a full complement of high performance peripherals employing standardized interfaces.

The following pages describe the Sun StorEdge A5000 architecture in detail. The entire system, from host adapter connections to interface boards, gigabit interface converters, the front panel, and high performance, standard FC-AL disk drives are discussed. Because the Sun StorEdge A5000 is supported by some very special software capabilities, a separate chapter describing the volume management, array management, and monitoring software immediately follows this one.

Capacity

To meet the wide-ranging capacity requirements of today's organizations, the Sun StorEdge A5000 is packaged in an enclosure suitable for both office and datacenter environments (Figure 3-1). Available in both deskside and cabinet configurations, the Sun StorEdge A5000 provides capacities ranging from 45 GB to over 12 TB, with higher capacities anticipated in the near future.

- The Sun StorEdge A5000 provides up to 127 GB of storage capacity in less than 2 cubic feet (0.06 cubic meters). It holds any combination of five to fourteen 9.1 GB, 7200 RPM, half-height 3.5 inch FC-AL disk drives (or up to 22 one-inch, 4.3 GB disk drives).
- Up to four Sun StorEdge A5000 arrays and two FC-AL hubs can be combined for a total capacity of 509 GB in a single datacenter cabinet. Multiple cabinets can be combined to provide over 12 TB of storage to users.

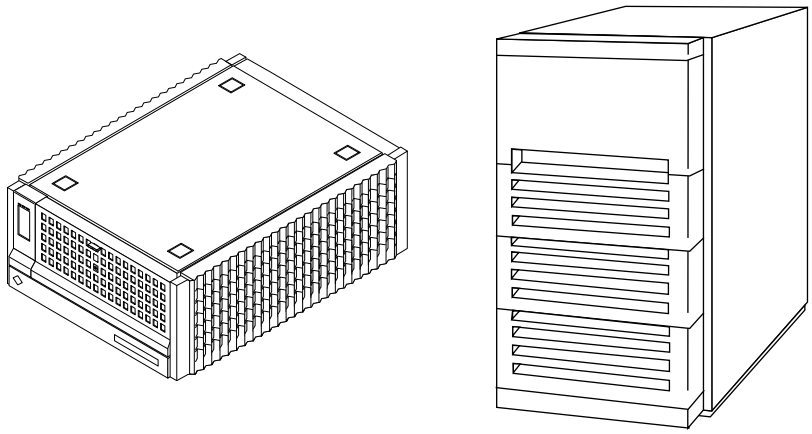

Figure 3-1 The Sun StorEdge A5000 can be used in tabletop and cabinet configurations

System Overview

The Sun StorEdge A5000 is a high availability mass storage subsystem that uses a disk enclosure capable of supporting up to 127 GB of storage with greater capacities expected in the near future. The system includes one or two interface boards, FC-AL drives, a front panel module, and one or more Fibre Channel Arbitrated Loop host adapters and optical modules for connection to a host system (Figure 3-2). Active components in the disk enclosure are fully redundant and may be replaced while the array continues to operate.

Key components of the Sun StorEdge A5000 include:

- Hot-pluggable Front Panel Module for configuration control and status
- FC-AL host adapter — SBus™ card with two independent FC-AL interfaces
- Dual hot-pluggable interface boards with 2 independent FC-AL sockets
- Front and rear disk drive backplanes, each supporting up to 7 (or 11) drives
- Dual-ported, hot-pluggable FC-AL disk drives
- Hot-pluggable FC-AL Gigabit (optical/electrical) Interface Converters
- Hot-pluggable fan trays: one in the front of the enclosure, one in the rear
- Three hot-pluggable power supplies
- No internal cables

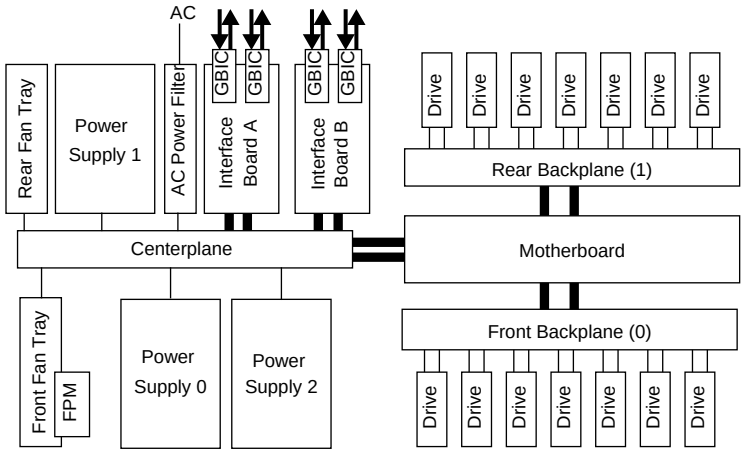

Figure 3-2 The Sun StorEdge A5000 system overview

Fibre Channel Interface

The Sun StorEdge A5000 is equipped with dual industry standard Fibre Channel Arbitrated Loop interfaces, each capable of running at 100 MB/second full duplex. Key features of the interface include:

- Dual-porting enables four independent Fibre Channel connections.
- Sun's SBus Fibre Channel Arbitrated Loop host adapter is also dual-ported and upgradable, enabling two loops to be connected to each host bus adapter. Using full cabinet configurations, up to 1 TB of high performance, high RAS storage can be connected to a single SBus slot.

- Because the Fibre Channel Arbitrated Loop interface uses 50 micrometer optical fiber, storage can be separated from the host by up to 500 meters (0.3 miles). The use of optical fiber also assures total electrical isolation from the host system. The use of single mode fiber permits distances up to 10 kilometers.
- The Fibre Channel interface on the Sun StorEdge A5000 is fully compliant with profiles recently released by the FC-AL standards committees to ensure interoperability among Fibre Channel-based products from major vendors. All FC-AL profiles are compliant with the ANSI Fibre Channel standards.

To provide optimum performance and to keep costs low, Sun developed a second generation Serial Optical Channel Plus (SOC+) FC-AL controller ASIC that can service two full-duplex, SBus to Fibre Channel conversions at 100 MB/second. Internally buffered, microprogrammable, and running downloadable firmware, the SOC+ matches the transfer rate of the SBus.

Interfacing the SOC+ to the Fibre Channel media is a Gigabit Interface Converter (GBIC). Each GBIC provides an interface to two 50/125 multimode fibers. (Two fibers are required.) Currently running at 100 MB/second, the GBIC cards support Sun StorEdge A5000 scalability and performance requirements. One GBIC is included with each Sun StorEdge A5000 host bus adapter, and an additional one can be added to support both Fibre Channel ports.

Gigabit Interface Converter Module

The Gigabit Interface Converter Module (GBIC) is a small, hot-pluggable optical/electrical conversion unit that converts standard Fibre Channel connector and signalling technologies to a standard copper serial connection. The GBIC operates at 1.0625 Gbits/second.

The standard GBIC included with the Sun StorEdge A5000 operates with optical connections. On one end is an electrical connection which interfaces with internal buses. The other end is an optical connection with a standard SC fiber cable connector (Figure 3-3). The GBIC uses a shortwave laser that operates at a low power level, precluding the need for Open Fiber Control safety circuits.

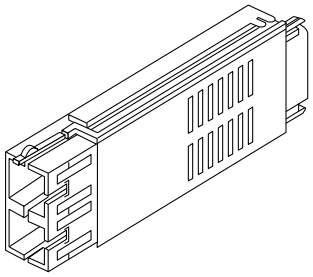

Figure 3-3 The GBIC used in the Sun StorEdge A5000 is a hot-pluggable unit that converts standard Fibre Channel to a standard serial connection

Controls to the GBIC allow for turning the transmitter on and off. Sense information from the GBIC indicates transmitter faults and loss of signal.

Using 50-micron fiber, host-to-array connections can extend to 500 meters. Using longwave lasers and single mode fiber cables, the Fibre Channel standard supports up to 10 km distances.

Interface Board

The Interface Board (IB) provides FC-AL connections to the enclosure. Furnishing all intelligent controls for the array, the IB supplies special services to report and control the state of the enclosure and its components, sensing and setting the environmental service signals as required by conditions inside the unit. The IB interprets enclosure service commands from the host or the front panel module and performs the indicated enclosure management and sensing functions.

The Sun StorEdge A5000 is designed to incorporate two identical, redundant interface boards in the lower rear of the enclosure. Each interface board serves one port on each FC-AL disk drive: interface board A serves port A on each FC-AL disk drive, and interface board B serves port B on each FC-AL disk drive.

The IB provides bypass services for two independent FC-AL ports and manages the configuration of the internal loops. The Sun StorEdge A5000 can be configured as a single loop, a dual loop, or a split loop.

- When operating as a single loop, port 0 or 1 of the interface board is connected to the Fibre Channel link. The other port is available for attachment of another Sun StorEdge A5000 enclosure or a host.
- When operating as a dual loop, port 0 of each FC-AL drive is connected to the top interface board, and port 1 of each FC-AL drive is connected to port 1 of the bottom interface board.
- When operating as a split loop, the front disk drives are connected to port 0 of the interface board and the rear disk drives are connected to port 1 of the interface board. This creates two separate FC-AL loops that can be connected to the host.

The Sun StorEdge A5000 uses a series of bypass circuits and multiplexors to enable reconfiguration of the nodes within the enclosure. Failing or missing devices and circuits may be bypassed, ensuring the remainder of the system remains available and service is not disrupted by the loss of a single component.

Drive Technology

One of the cornerstones of disk array technology is its use of commonly available, highly reliable, low cost disk drives. By exploiting these mass produced drives, cost-effective products like the Sun StorEdge A5000 are made possible.

FC-AL Interface

All drives in the Sun StorEdge A5000 contain an FC-AL interface that supports the SCSI command set. Each drive uses a small form-factor, 40 pin single connector attachment (SCA) for FC-AL interconnect. The 3.5” FC-AL disk drives are dual-ported, operate at 7200 RPM and are 1.6” high (half-height).

Key functions of the FC-AL drives include:

- The FC-AL drives are dual-ported for multi-path access. The A and B disk ports can be accessed through separate and independent interface boards.
- Each FC-AL disk drive is connected directly to the Fibre Channel loop and appears as a node on the loop. Each drive is identified by a World Wide Name that is unique.

- The FC-AL drives accept all standard SCSI-3 commands. Since SCSI commands are delivered directly to the drives via the Fibre Channel loop, a legacy copper SCSI interface is not required, ensuring higher throughput and lower latency.
- FC-AL drives are hot-pluggable. Software preparations, however, must be made for removal, replacement, and additions to be properly recognized and configured.

Single Connector Attachment System

Single Connector Attachment (SCA) is a technology jointly developed by Sun and drive manufacturers that incorporates all electrical and signal connections into a single, blind-mating connector, enabling drives to be directly plugged in to a compatible disk tray, such as those used in the Sun StorEdge A5000. Unlike attempts by others to retrofit existing drives using carriers and unreliable flex cables, jumpers and wires, SCA is built directly into the drives. All drives in the Sun StorEdge A5000 use a standard 40 pin SCA for FC-AL interface.

Drives using Single Connector technology feature higher reliability, better customer serviceability, and lower cost. Multiple manufacturers produce drives compatible with the Single Connector system, and Single Connector technology is a freely available industry standard.

3.5″ Drives

An important decision in the design of the Sun StorEdge A5000 is the use of half-height 3.5″ drives. While Sun engineers could have chosen larger, higher capacity drives, they knew that an absolutely key determinant of random I/O performance is a high ratio of spindles to total capacity. Using the thinner drive allows a greater number of drives in the array, permitting a better balance of performance, flexibility, and capacity. Up to 14 half-height drives (or up to 22 one-inch drives) can be accommodated in a single Sun StorEdge A5000 enclosure.

Host Adapter

The Sun StorEdge A5000 employs a single-width Fibre Channel SBus card to communicate with the host. The SOC+ Host Adapter (SOC+HA) operates in 32-bit or 64-bit mode, and has a second generation Sun Serial Optical Channel

Plus (SOC+) ASIC processor on-board. The host adapter contains two independent FC-AL interfaces, each operating at 100 MB/second (Figure 3-4). One or two FC-AL loops can be connected to each host adapter using hot-pluggable Gigabit Interface Converters.

The SOC+HA card supports both FC-AL loop and point-to-point connections. The card provides an open interface for connecting other devices that meet the same Fibre Channel Protocol standards.

The host command buffer (HCB) and the SOC+ programming interface process requests within a single interrupt. Tagged queueing is supported, and multiple response entries may be in the queue when the host services the interrupt, making it possible to achieve less than one interrupt per I/O request and increasing overall perceived throughput.

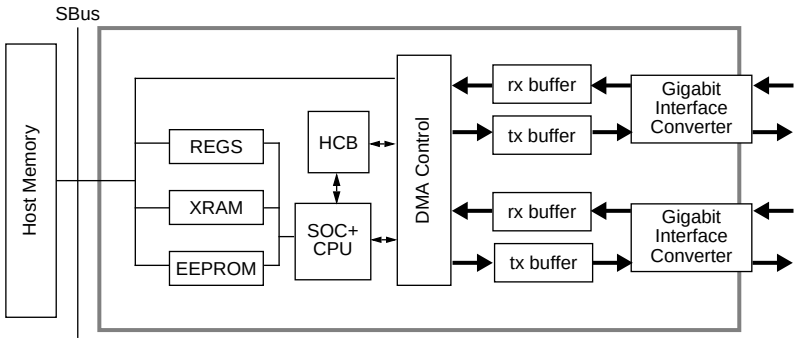

Figure 3-4 The FC-AL host adapter functional block diagram

Enclosure Services and the Front Panel Module

Designed to provide high levels of serviceability, the Sun StorEdge A5000 supports SCSI Enclosure Services (SES) — a device model that enables configuration and maintenance information, including status, to be displayed locally and conveyed to a host system. The SCSI Enclosure Services run on the SOC+ chip on the designated interface board and communicate via the SCSI-FCP protocol across the FC-AL interface.

All enclosure services are performed by the processor on the SOC+ chip on the appropriate interface board. If only one interface board is installed and operational, that interface board performs all the enclosure services. If two interface boards are present, enclosure service tasks are divided among the two cooperative boards. Status, warning, and critical fault conditions are available for all monitored components in the array, including the power supplies, fan trays, interface boards, GBICs, disk drives, and disk drive backplanes.

Front Panel Module

The Front Panel Module enables an operator to interact with the Sun StorEdge A5000 disk subsystem. The FPM accepts touch switch inputs and provides graphic and alphanumeric information on an electro-luminescent display (Figure 3-5). Three LEDs provide additional summary status information.

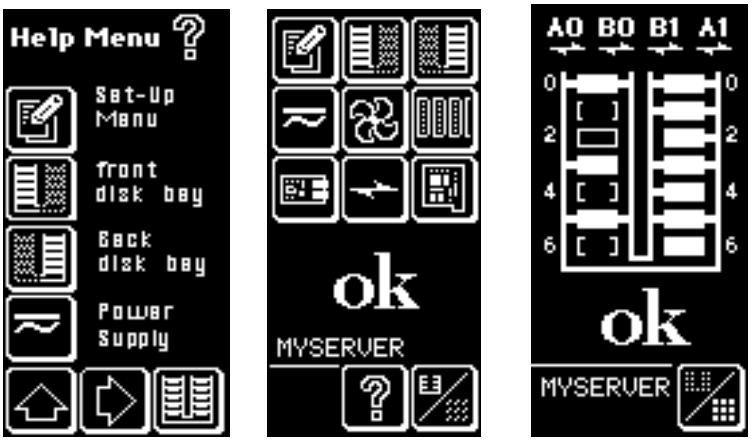

Figure 3-5 The Sun StorEdge A5000 Front Panel Module

The FPM has three main functions:

- Displays enclosure, drive, and loop status and highlights errors
- Displays vital enclosure data, including worldwide name, box name, and box ID
- Configures the enclosure, setting box name and loop configuration

The FPM touch screen is a 3x6 array of touch areas which are under-labelled by images from the graphic display, indicating active/passive state and what actions are performed by each button. The touch screen also provides numeric inputs to the enclosure and provides eighteen buttons that enable a system administrator to step through the diagnostic and display menus.

The FPM supplements the enclosure services provided through the SCSI-3 Environmental Services command set. The FPM provides access to these and additional services even if the FC-AL is not connected or if the host processor's monitor and keyboard are not at the same physical location as the array.

Enclosure and Power

Packaging is frequently considered as an afterthought in many disk array products, but Sun understood that properly designed, packaging could make a significant contribution to the reliability and serviceability of the Sun StorEdge A5000. Elegantly packaged, drives can be added or replaced in just one minute without configuring jumpers or shutting down the system. All components are interlocked to prevent damage or exposure to dangerous voltages. Disk drives employ safety catches to prevent them from being dropped accidentally.

The Sun StorEdge A5000 can be placed horizontally on a tabletop, or multiple units can be placed inside a standard cabinet for datacenter applications (Figure 3-6).

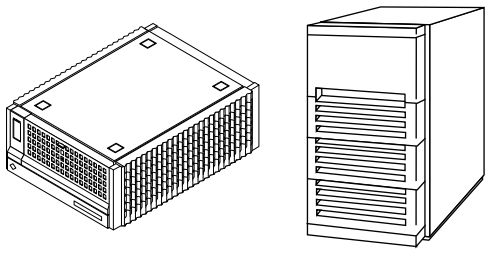

Figure 3-6 The Sun StorEdge A5000 can be placed horizontally on a tabletop, or in a standard datacenter cabinet

The Sun StorEdge A5000 can be used in both tabletop and cabinet configurations. All configurations share the same architecture, electronics, and volume management software. The difference is in the packaging, maximum capacity, and suitability for certain tasks.

- The Sun StorEdge A5000 is compact and easily serviceable. Designed to fit in a 20″ wide by 25″ deep by 9″ high tabletop cabinet, the enclosure is ideal for a wide range of environments, from small workgroups to departmental applications with storage requirements in the 45 GB to 127 GB range.
- Multiple Sun StorEdge A5000 units can be grouped in a cabinet, enabling organizations to build very large storage systems geared toward supporting large server systems like Sun's Enterprise X000 product family. Combining up to four Sun StorEdge A5000 units and FC-AL hubs, the subsystem can accommodate 509 GB in a single datacenter cabinet.

Key Components

All components in the Sun StorEdge A5000, including drives, fans, power supply, and interface boards, are plugged into special front and rear backplanes that connect all power, signal, and data lines without using cables (Figure 3-7).

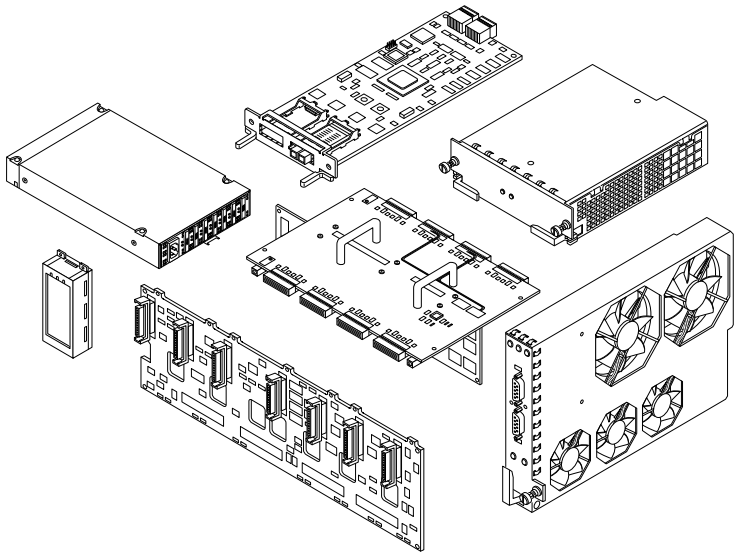

Figure 3-7 The Sun StorEdge A5000 consists of modular components that interconnect through front and rear backplanes

Three redundant, hot-swappable and auto-ranging power supplies provide power from 100-240 VAC and 47-63 Hz. Easily replaced in the event of failure, their 200,000 hour MTBF at 35˚C rating ensures high reliability. The array subsystem can remain operational with only two power supplies.

Fan Trays

The Sun StorEdge A5000 is cooled by two independent, thermostatically controlled fans. One fan tray is in the front of the unit, and the second is located in the rear of the unit. All fan trays in the array enclosure are hot-swappable, and the system can remain operational with a single fan failure.

If a fan fails, both the front panel module and array manager software will alert administrators, giving them ample time to organize replacement without interrupting operation. Only when an actual over-temperature condition occurs will the system automatically shutdown. As with other components in the Sun StorEdge A5000, the fan units are hot-swappable and can be quickly replaced without tools or special training.

The fan trays are monitored components. The FPM and SCSI Enclosure Services can report the presence and revision levels of fan trays, as well as single and multiple fan failures.

FC-AL Hubs

Organizations that have large amounts of data, or anticipate storage growth beyond a single Sun StorEdge A5000 can choose how to connect that storage to more than one host system. Arrays can be daisy-chained together or connected to a hub (Figure 3-8). While daisy-chaining has traditionally been used as a low cost mechanism to extend cable length and tie subsystems together, it is complex, with awkward cabling and limited flexibility.

The use of FC-AL hubs mitigates the drawbacks typically associated with device daisy-chaining. FC-AL hubs connect attached devices to a single loop. The hub becomes, in essence, a loop in a box — everything, both arrays and hosts, plugged into the hub becomes part of the same loop. When used with the Sun StorEdge A5000, FC-AL hubs offer a host of benefits:

- Reduced wiring complexity
- Rapid insertion of additional arrays or hosts
- Automatic bypass of arrays which are not operational, preventing failure of the remainder of the loop

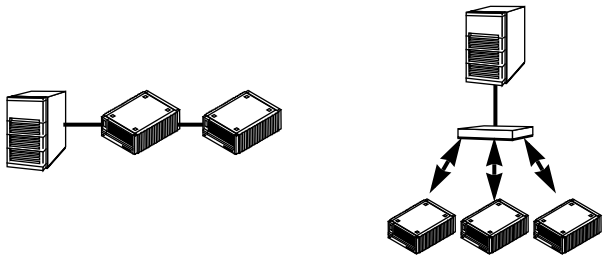

Figure 3-8 Multiple Sun StorEdge A5000 units can be daisy-chained together or connected via FC-AL hubs

The Sun StorEdge A5000 FC-AL hub has a number of characteristics that enable it to increase data availability and simplify the cabling of multiple arrays:

- Seven FC-AL ports
- Holds up to seven GBIC optical modules
- Automatically reconfigures, bypassing failed components
- Supports up to one loop per hub
- Up to two hubs can be housed in a Sun StorEdge A5000 cabinet

Designed for Reliability, Availability, and Serviceability

The Sun StorEdge A5000 was designed specifically to meet the reliability, availability, and serviceability requirements of today's organizations. Utilizing redundant, hot-swappable components, Fibre Channel technology, various RAID levels, and sophisticated volume management and configuration software, the Sun StorEdge A5000 includes a host of RAS features to support the most demanding environments.

Reliability

A number of factors affect the reliability of a mass storage system, from individual component quality, to environmental conditions, to data reliability under nominal and exception conditions.

- To ensure high uptime, each component in the Sun StorEdge A5000 has undergone extreme quality testing, ensuring they meet the most rigorous standards.

- All components in the Sun StorEdge A5000 have very high individual MTBFs and MTTDL, contributing to overall subsystem uptime in excess of 99.9%.
- All drives, power supplies, fan modules, and cables in the array are hot-pluggable and redundant, providing both reliability and availability in a flexible design.
- The array is equipped with temperature monitoring sensors to warn of impending cooling problems and to automatically initiate a controlled shutdown in the event predefined limits are exceeded.

Availability

High performance alone is not enough — modern applications also require increased uptime and availability. With requests for data able to originate from anywhere in the world, any time of day, it is important that data availability is not compromised by hardware failure. Storage systems can only fail infrequently, and when they do, problem diagnosis and repair must be swift.

The Sun StorEdge A5000 has a variety of features to ensure its availability and uptime:

- Beginning with highly reliable individual components, Sun placed them in innovative packaging entirely free of internal cabling. Using redundant, hot-swappable modules, all components in the array can be quickly and easily replaced by a system administrator or operator. If dual loop configurations are employed, failed loops and devices can be bypassed without operator intervention.
- Multi-pathing is inherent in the Sun StorEdge A5000 design. Dual-ported drives, multiple loops configurations, and dual connections to the host ensure a high degree of device path availability. Combined with Dynamic Multi-Pathing support, the Sun StorEdge A5000 provides the data availability users demand even in the event of a device path failure.
- Drives in the array can be designated as hot spares, enabling them to take over for any failed drive in a RAID 1, 0+1, or 5 configuration. Error detection, recovery, and restoration to the designated spare is automatic by default.

- Hot swap of a failed drive is supported; the replacement drive is detected automatically in the enclosure, and restoration of the new drive (and return of the spare drive) is managed from the host software.
- Using RAID, the Sun StorEdge A5000 is capable of over 99.9% uptime. Simultaneously able to support any combination of RAID levels 0, 1, 0+1, 5, and independent spindles, the Sun StorEdge A5000 enables administrators to carefully match their reliability goals with the right configurations, including high availability system fail-over arrangements. Indeed, the Solaris™ Cluster platform, Sun's latest high performance, high reliability clustered environment, incorporates the Sun StorEdge A5000 as an integral part of its configuration (Figure 3-9).

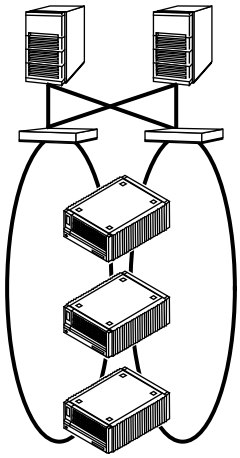

Figure 3-9 Dual-ported Sun StorEdge A5000 units are well suited for incorporation into high availability server configurations

Serviceability

While Sun was able to exploit many technologies already proven in earlier disk arrays, workstation, and server products, when they created the Sun StorEdge A5000 they knew that additional developments were also required. In particular, innovations were needed to increase serviceability for mission-critical applications. The Sun StorEdge A5000 is compelling for just this reason — it mixes proven reliable components with a subsystem hardware and software architecture that enhances serviceability.

- By simplifying system administration and configuration tasks, the Sun StorEdge A5000 promises to lower operating costs.
- The cable free, hot-plug design of the array enclosure and SCA-2 disk drives ensure complete customer serviceability.
- Identifying and responding quickly to bottlenecks caused by changing conditions requires tools that support rapid system reconfiguration. The array's Front Panel Module, *luxadm* configuration tool, and the Sun Enterprise Volume Manager™ software aid in this effort through their ability to support downloading of firmware, FC-AL loop configuration, and subsystem performance monitoring.
- All disk array operations can be controlled through a set of shell commands or through a set of library routines which administrators can use to develop special maintenance scripts or programs.
- Pinpointing and diagnosing disk and disk array status is critical to servicing the subsystem. The Sun StorEdge A5000 has been integrated with SunVTS™ and Solstice™ SyMON™ — a robust, comprehensive set of tools that perform runtime diagnostics, fault isolation, and performance monitoring, and aid failure prevention.

Performance

The Sun StorEdge A5000 excels in both transaction and transfer rate performance, and incorporates RAID technology to balance system performance and availability.

- The subsystem loop is capable of over 10,000 uncached two-KB input-output operations per second (IOPS), and sustained transfer rates exceeding 95 MB/second — enough to meet the needs of the most demanding applications.
- Peak bandwidth of 190 MB/second is achieved in dual-loop configurations performing 1 MB sequential reads.
- The Sun StorEdge A5000 can simultaneously support multiple RAID levels, including level 0 (striping), 1 (mirroring), 5 (striping with distributed parity) and 0+1 (striped and mirrored), in addition to independent spindles — techniques which can improve read/write transfer rate performance and data availability.

Scalability

Businesses appreciate products that have long and useful lives. Rather than having to replace equipment to get more performance or capacity, upgrading existing components is far more attractive. Upgrades usually represent incremental costs, often falling within discretionary budgets. Extensive retraining or reconfiguration is rarely needed, and compatibility problems are infrequent.

A scalable storage system must be physically and architecturally capable of growing in capacity and performance. The Sun StorEdge A5000, with its ability to deliver over 10,000 uncached I/O operations per second, and over 95 MB/second of sustained transfer rate can easily meet the requirements of most applications. For those that require more, multiple Sun StorEdge A5000 units can be combined. Table 3-1 shows the maximum tested and supported storage capacity of Sun's server systems equipped with Sun StorEdge A5000 subsystems.

	Maximum Sun StorEdge A5000 127 GB Arrays	Maximum Sun StorEdge A5000 509 GB Arrays	Maximum Total Capacity
Sun Enterprise Server 3000	24	6	3 TB
Sun Enterprise Server 4000	48	12	6 TB
Sun Enterprise Server 5000	48	12	6 TB
Sun Enterprise Server 6000	72	18	9 TB
Sun Enterprise Server 10000	—	24	12 TB
SPARCserver™ 2000/2000E	72	18	9 TB
SPARCserver™ 1000/1000E	8	2	1 TB

Table 3-1 With the Sun StorEdge A5000, Sun servers can deliver very large on-line storage capacities

With dual 100 MB/second, full duplex ports, the Fibre Channel Arbitrated Loop interface of the Sun StorEdge A5000 ensures that there are no bottlenecks in the link between the host and the storage system. Fibre Channel's ability to connect with high speed switched communications "fabrics" enables today's sophisticated Sun StorEdge A5000 to become part of very large intelligent storage network systems.

3

Specifications and Standards

The Sun StorEdge A5000 meets all relevant domestic and international agency safety, ergonomics, EMI, and environmental requirements.

The Sun StorEdge A5000 also conforms to all related industry standards, including:

- Fibre Channel Physical and Signaling Interface (FC-PH) ANS X3.230-1995
- Fibre Channel Physical and Signaling Interface (FC-PH-2) ANS X3.297-1997
- Fibre Channel Physical and Signaling Interface (FC-PH-3) ANS X3.303-1997
- Fibre Channel Arbitrated Loop (FC-AL) ANS X3.272-1996
- Fibre Channel Private Loop, SCSI Direct Attach (FC-PLDA), Project 1162-DT, revision 2.10
- Fibre Channel Protocol for SCSI, ANS X3.269-1996
- SCSI-3 Architecture Model (SAM), ANS X3.270-1996
- SCSI-3 Primary Commands (SPC), ANS X3.301-1997
- SCSI-3 Block Commands (SBC), ANS NCITS.306-1997
- SCSI-3 Enclosure Services, ANS NCITS.305-1997
- 40-Pin SCA-2 Connector w/ Bidirectional ESI, revision 1.8 or later, Small Form Factor Committee SFF-8067
- SCA-2 Unshielded Connections, Small Form Factor Committee SFF-8451, revision 8.0 or later

Sun StorEdge A5000 Software 4

Disk arrays have managed to earn a reputation for high performance and reliability, but generally at the cost of flexibility and administration ease. Reconfiguring arrays in response to new requirements or system faults is often complex and time-consuming. Understanding this, Sun was determined to make the management of the Sun StorEdge A5000 more straightforward than competing products, while at the same time using array configuration, management, and monitoring software to further increase the reliability, serviceability, capacity, and performance of the entire system.

Sun Enterprise Volume Manager™

Sun Enterprise Volume Manager is a software management solution that helps users manage disks and disk arrays connected to Sun systems, enabling users to more easily address their storage management, administration, availability, and performance needs:

- The software brings disk storage management under a uniform administrative system that operates while volumes are on-line and being accessed by users and applications.
- The software improves data availability by enabling administrators to easily create consistent copies, or mirrors, of stored data.
- Sun Enterprise Volume Manager assists in improving array performance by permitting the creation of striped volumes, and works with the special capabilities of the Sun StorEdge A5000 to reduce the burden of certain RAID-specific operations on the host system.

- Sun Enterprise Volume Manager provides mechanisms that enable the monitoring and tuning of array performance. Tuning is permitted while the array continues to operate.
- The architecture of the Sun Enterprise Volume Manager easily allows the incorporation of new capabilities while retaining compatibility with existing applications.
- Sun Enterprise Volume Manager reduces the cost of disk management under Solaris by supporting dynamic repartitioning and on-line backup. The ability to perform routine maintenance chores while the system is being used minimizes the need for additional personnel and provides greater overall uptime.

Administrators configure and control Sun StorEdge A5000 arrays using Sun Enterprise Volume Manager. With a graphical interface that includes pull down and cascading menus, forms, and dialogue boxes, the Volume Manager also includes an extensive context-sensitive help facility to assist users.

With the Volume Manager, it is easy to quickly identify and eliminate disk I/O bottlenecks. Forms can be displayed which show the total number of I/O operations, percentages of read and writes, and total blocks read. Volume traffic can also be indicated by color, green being least busy, yellow warning of increasing activity, and red indicating a possible performance bottleneck. The Volume Manager software enables administrators to identify and correct problems on-line without disrupting the work of users or applications, resulting in greater system availability and better overall throughput. Hot spots can be eliminated by simply dragging them closer together to reduce seek latency, moving a them to other, less-used disks, or adjusting the organization of a RAID group.

The Sun Enterprise Volume Manager also simplifies and streamlines routine tasks such as backup. Using the snapshot operation to create a mirror of an existing volume, a consistent copy of an original volume can be made at the time the snapshot was begun. The snapshot volume can then be used to make a backup of the original volume without interrupting access to it. Once the backup is completed, the snapshot volume can be removed.

In order to provide a single, consistent interface for disk administration, the Volume Manager can be used to control all disks connected to a system, including those not contained in a Sun StorEdge A5000 enclosure. The only restrictions are that there must be at least one Sun StorEdge A5000 on the

system, and that striped data configurations (RAID 0, 0+1, and 5) cannot be comprised of a combination of Sun StorEdge A5000 disks and other external disk drives.

Using Volume Manager Volumes

Once volumes have been created, they can be used in the same manner as physical partitions. The Volume Manager hides all details of disk configuration and RAID groups from applications, databases, and file systems. Regardless of the underlying configuration, volumes appear as partitions. This means that applications can exploit the performance, capacity, and reliability of the Sun StorEdge A5000 without modification. File systems can be created on them, or they may be used as raw devices by applications such as database management systems. The Volume Manager even allows the UNIX® root partition to be placed on a volume. Once created, volumes can be reorganized whenever necessary. This "on-the-fly" reorganization enables administrators to change volume or file system size, add mirrors, and perform other tasks without interrupting users.

One of the principal tasks of the Volume Manager software is to ensure recovery in the event of a system failure. Because all Volume Manager I/O operations occur atomically, operations wholly succeed or wholly fail, ensuring that data is not left in an indeterminate state. This means that any recovery activities, such as mirror resynchronization, can take place after the system is back on-line, ensuring a quick return to normal operation.

Volume Manager Architecture

The device driver architecture has been modified to accommodate the Sun StorEdge A5000. Its components include:

- *Dynamic Multi-Pathing (DMP) Driver*

  A high-level device driver that automatically recognizes multiple I/O paths to a particular disk and provides path failover in the event of a lost connection.

- *SCSI Disk Driver*

  The standard Solaris SCSI disk driver.

- *Nexus Controller*

  The Nexus Controller provides an interface between the controller interface library routines and the lower-level routine responsible for preparing commands for the array.

- *SCSI-Fibre Channel Nexus Driver*

  This portion of the device driver chain forms outbound SCSI command descriptors into packets that are compatible with the SCSI-FCP protocol. It also produces SCSI-compatible descriptors and result codes from inbound blocks conforming to the SCSI-FCP format.

- *Fibre Channel (SOC) Driver*

  This lowest layer of the driver provides a generic Fibre Channel interface and is responsible for register-level control of the Serial Optical Channel.

Command Line and Programming Interface

In addition to the graphical interface, the Volume Manager and Sun StorEdge A5000 interface software includes a complete set of scriptable utilities that can be used to set up volumes, create disk groups, add file systems to volumes, add and remove mirrors, and perform special service-related administrative tasks.

luxadm

The Sun StorEdge A5000 includes an administrative command (luxadm) to help configure and administer the array. A second-generation utility, luxadm enables the system administrator to:

- Probe the configuration for all attached FC-AL array devices
- Display information about attached enclosures and disk drives
- Download firmware
- Control configurations and operating options
- Manage power requirements
- Employ alarm functions
- Prepare the array for device removal or insertion
- Power up/down and enable/disable components
- Operate the array remotely

Solstice™ SyMON™

A new force contributing to change in client-server computing is the use of mission-critical applications across enterprise-wide networks. As a result, distributed computing environments must be more reliable, available around the clock, and easier to diagnose and service. Solstice SyMON is a powerful tool for monitoring the operation of the intelligent storage network. It enables system administrators to perform remote server and peripheral configuration, monitor performance, and isolate hardware faults all through an easy-to-use graphical user interface.

Today, Solstice SyMON extends its rich set of features by supporting the Sun StorEdge A5000. With Solstice SyMON users can:

- Display physical and logical views of the Sun StorEdge A5000
- Display configuration information
- Display component-specific information
- View FC-AL disk drives in the enclosure
- Obtain FC-AL disk drive status
- Generate alerts for over temperature and under/over voltage conditions
- Test power supplies, interface boards, and GBICs
- Check backplane and EPROM status

Furthermore, users can configure Solstice SyMON to initiate actions when specified event occur, including SNMP trap generation, administrator paging, phoning home, and more.

SunVTS™

Although much of today's computer hardware is designed with high reliability in mind, failures do occur — and users need a robust tool that can help detect, locate, and isolate system and component failures. SunVTS is a graphical diagnostic tool designed to help monitor and diagnose system problems. With SunVTS, system administrators and users alike can perform runtime diagnostics, isolate faults, and obtain status information on all devices, enabling the swift resolution of failures.

With SunVTS, users can validate, exercise, and isolate faults in their storage networks. The latest release of SunVTS has been enhanced to include support for the Sun StorEdge A5000. With SunVTS 2.1 users can:

- Check Sun StorEdge A5000 status
- Determine the status of power supplies fan trays

- Obtain the status and configuration of FC-AL drives and loops
- Detect over temperature conditions
- Check the status of the Fibre Channel interface and related components
- Test the FC-AL host adapter
- Check the paths between a host and a Sun StorEdge A5000 enclosure
- Exercise the FC-AL disk drives and data paths
- Isolate component failures, including GBIC, backplane, disk, host adapter, loops

Application Environments 5

The Sun StorEdge A5000 solves important problems in a variety of new and growing application areas. Intended to address obstacles frequently encountered when configuring high performance servers and workstations, the Sun StorEdge A5000 challenges many traditional assumptions about high performance mass storage systems and provides unique solutions to a variety of environments characteristic of key enterprise applications.

On-line Transaction and Database Processing

Nearly every database management problem has its unique requirements — some need high throughput for large records, others need rapid response in transaction-oriented applications. As enterprises become more service oriented, reliability and uptime is becoming an urgent requirement as well.

The Sun StorEdge A5000 is positioned to deliver any needed mix of performance and reliability. By supporting multiple RAID levels and independent spindles, administrators have the flexibility to balance their requirements with budgetary and practical constraints. The Sun StorEdge A5000's ability to maintain hot spares means that neither redundancy or performance will be compromised in mission-critical database environments.

OLTP environments are characterized by small, random reads and writes. Typically not a bandwidth-intensive application, OLTP places its most stringent demands on address availability — devices and the data on them are

critical. The Sun StorEdge A5000 strengthens OLTP environments with its support for dual paths between the host and array and the ability to mirror critical data (Figure 5-1).

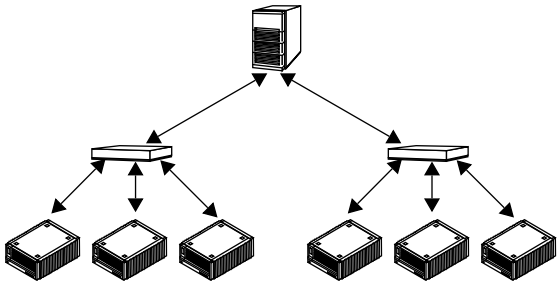

Figure 5-1 The Sun StorEdge A5000 supports OLTP environments

Data Warehousing and Decision Support

The key to a successful data warehouse is to start small, but plan big. Starting small, using iterative delivery techniques, allows a rapid, consistent demonstration of business value — without getting caught up in an all-or-nothing warehouse implementation. Growth is inevitable due to scale-ups in data, users, and workloads. Users and IT professionals alike inevitably underestimate the growth and processing requirements for a successful data warehouse. By the time the database is constructed with indices and a modest number of summary tables, required storage capacity is at least three times the raw data size. That factor can easily grow much larger if numerous summary tables and indices are created for performance enhancement.

A critical factor to the success of a data warehouse or decision support system is raw capacity — the ability to access hundreds of terabytes of information in a timely manner is key. Characterized by large sequential read requests, data warehousing and decision support systems require large storage capacity and high bandwidth. The ability to employ dual FC-AL paths between hosts and arrays enables the Sun StorEdge A5000 to meet not only the performance needs of these environments but availability requirements as well.

Network File Service

A large number of server systems are dedicated to providing file service to networks of distributed workstations and personal computers. Environments using network file systems, such as NFS™, can be described in one of two ways:

- *attribute intensive*, characterized by bursts of short transactions, or
- *data intensive*, where transactions are less frequent, but longer.

The Sun StorEdge A5000 offers a powerful, scalable platform to address both of these problems. The Sun Enterprise Volume Manager software also ensures that administrators can quickly respond to changing conditions created by new applications and user loads, ensuring the Sun StorEdge A5000's effectiveness for environments needing high performance file service.

In network file server environments — environments characterized by small, random reads and writes — bandwidth is generally not the chief concern. What is critical, however, is the availability of the server and data it provides. Utilizing dual paths between the host and array, as well as mirroring techniques, the Sun StorEdge A5000 provides the availability NFS environments require. Larger systems, or those which handle large I/O requests for data-intensive applications like seismic analysis, the use of dual-pathed 100 MB/second Fibre Channel Arbitrated Loop technology and striping techniques are critical to providing the increased performance these specialized environments demand.

Imaging

Imaging is a broad problem that encompasses two kinds of applications: 1) the electronic storage of printed documents for later retrieval and reference, and 2) the digital representation of images—pictures—for use in publishing, entertainment, archives, engineering, or science. Both are characterized by very large file sizes that can become extremely cumbersome to manage and which require I/O performance that many disk arrays cannot achieve.

With capacities measured in the hundreds of terabytes and transfer rates over 95 MB/second, the Sun StorEdge A5000 is ideally suited for use with high performance imaging workstations and servers. With its excellent scalability, additional disk storage can be easily added. The ability to daisy-chain arrays or connect them via FC-AL hubs permits multiple arrays to be configured in each

SBus slot, and the dual-pathed 100 MB/second Fibre Channel Arbitrated Loop interface and striping configurations deliver ample performance for the most demanding applications.

Interactive Services and Multimedia

The industry is abuzz with plans for a national network that will deliver a wide range of multimedia services directly to the home. Perhaps the most exciting of these will be video-on-demand, where future subscribers will be able to see movies, educational programs, and perhaps each other, whenever they want. Managing such video services will be a huge task for future servers. Systems will have to be able to sustain over 125 MB/second transfer rates just to meet the demands of 500 users, with a 2-hour movie requiring nearly two gigabytes of storage.

With appropriate allied technologies, a single Sun StorEdge A5000 cabinet can hold over 250 hours of broadcast-quality video and service the needs of multiple independent video “streams”. Multiple arrays, combined with the performance and scalability of multiprocessor SPARC-based servers, have the potential to deliver the performance needed for video editing and archiving, multimedia, and data collection.

Clusters

Clustered architectures are uniquely suited for the provision of highly available services. In a properly designed arrangement, they feature redundant paths between all systems, between all disk subsystems, and to all external networks. No single point of failure — hardware, software or network — can bring a cluster down. Fully integrated fault management software in the cluster detects failures and manages the recovery process without operator intervention, allowing failed components to be replaced on-line, without impacting availability. Such configurations promise the ability to provide the high levels of service needed without requiring high priced, proprietary technology.

The Sun StorEdge A5000 is uniquely designed to meet the challenges placed upon mass storage subsystems by clusters. Its redundant, hot-swappable components, support for RAID levels 1, 0+1, and 5, and dual paths between hosts and arrays enable the Sun StorEdge A5000 to provide unprecedented

levels of availability to clustered systems (Figure 5-2). Using hubs, cabling the Sun StorEdge A5000 in complex clustered environments is easily accomplished.

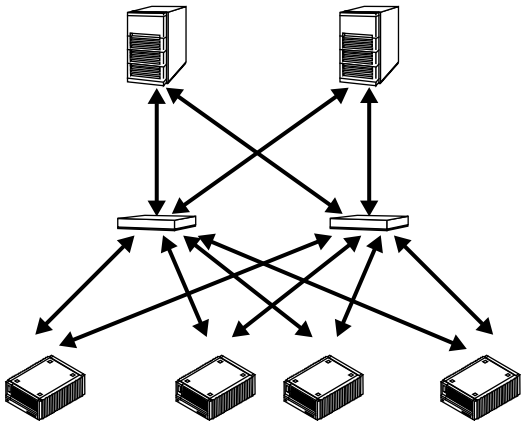

Figure 5-2 The Sun StorEdge A5000 supports clustered environments

5

Summary 6

Rarely do products so completely address user needs with so little compromise as does the Sun StorEdge A5000. Developed with a set of architectural features that make it one of the richest disk array products available anywhere, it incorporates state-of-the-art packaging, an intuitive front panel module, high-capacity drives, Fibre Channel Arbitrated Loop interface, and an advanced volume management utility — giving it the characteristics demanded by users and applications in mission-critical computing environments around the world:

- High capacity
- High performance
- High reliability
- High serviceability
- Hot plug capability
- Modular design
- Exceptional scalability
- Affordability
- Flexibility
- Low cost of administration
- Ease of use

Glossary A

Availability

A measure of the total time that data is available from a system.

Arbitrated loop

A loop technology in which two or more ports can be interconnected, but only two ports at a time may communicate.

Channel

An interface directed toward high speed transfer of large amounts of information.

Concatenation

A volume created by sequentially mapping blocks on disks to a logical device. Two or more partitions can be concatenated and accessed as a single device.

Disk array

A subsystem that contains multiple disk drives, designed to provide performance, high availability, serviceability, or other benefits.

Disk group

A grouping of disk drives and the data on them that facilitates organization and the movement of disks between systems.

Fabric

A group of interconnections between ports that includes a hub, switch, or arbitrated loop.

A

Fiber

A wire or optical strand. Spelled “fibre” in the Fibre Channel name.

Fiber-optic cable

Jacketed cable container of thin strands of glass through which pulses of light transmit data. Used for high speed transmission over medium to long distances.

Frame

An indivisible unit for transfer of information in Fibre Channel networks.

Full duplex

A communications protocol that permits simultaneous transmission in both directions, usually with flow control.

GBIC

Gigabit Interface Converter.

GUI

Graphical User Interface.

Hot relocation

A Sun Enterprise Volume Manager feature in which data is automatically reconstructed on a spare disk after a disk failure without interruption to user access.

Hot Spare

A drive in an array that is held in reserve to replace any other drive that fails. Hot spares are continuously powered up and spinning. This allows the array processor to have immediate access to a functioning drive for possible reconstruction of lost data.

Hub

A device used to connect fiber cables.

I/O rate

A measure of the capacity of a device to transfer data to and from another device within a given time period; typically stated in IO operations per second.

IOPS

Input/output operations per second. A measure of I/O performance, this is commonly used to quote random I/O performance.

IP

Internet Protocol. A set of protocols developed by the United States Department of Defense to communicate between dissimilar computers across networks.

Laser

Light Amplification by Stimulated Emission of Radiation. A device for generating coherent radiation in the visible, ultraviolet, and infrared portions of the electromagnetic spectrum. Used as an illumination source in fiber-based communications.

LED

Light Emitting Diode.

Link

One inbound fiber and one outbound fiber connected to a port.

Micron

One millionth of a meter. Also called micrometer.

Mirroring

In RAID terminology, refers to the redundant storage of data by duplicating the data.

Multimode fiber

An optical fiber which allows light to travel along multiple paths, resulting in gradual signal degradation with distance.

Network

An arrangement of nodes and connecting branches, or a configuration of data processing devices and software connected for information exchange.

N_Port

A port attached to a node for use with point-to-point or fabric topology.

NL_Port

A port attached to a node for use in all three topologies (point-to-point, arbitrated loop, or fabric).

Node

A device that has at least one N_Port or NL_Port.

Optical fiber

Any filament of fiber, made of dielectric material, that guides light.

Parity

In an array environment, data that is generated from user data and is used to regenerate user data lost due to a drive failure. Used in RAID 5.

Point-to-point

A topology in which exactly two ports communicate at a time.

Port

An access point in a device where a link attaches.

Protocol

A convention for data transmission that defines the sequencing of information transfers.

RAID

Redundant array of independent disks. A set of disk drives that appear to be a single logical disk drive to an application such as a database or file system. Different RAID levels provide different capacity, performance, availability, and cost characteristics.

RAID-0

RAID level 0, or striping. Data is distributed among disks for performance. No redundancy is provided, and the loss of a single disk causes the loss of data on all disks.

RAID 0+1

The combination of striping and mirroring. Data is distributed among disks for performance, and mirroring is used to provide redundancy.

RAID-1

RAID level 1, or mirroring. Multiple copies of the data are kept. This is inherently expensive because 100% duplication of data is required.

RAID-5

RAID level 5, or striping with distributed parity. Both data and parity are distributed across disks. No single disk can compromise the integrity of the data. RAID-5 optimizes performance, reliability and cost.

Receiver

The circuitry that receives signals on a fiber, and the ultimate destination of data transmission.

Redundancy

Duplication for the purpose of achieving fault tolerance. Refers to duplication or addition of components, data and functions within the array.

Responder

The logical function in an N_Port responsible for supporting the exchange initiated by the originator in another N_Port.

SCSI

Small Computer Systems Interface. An ANSI standard for controlling peripheral devices by one or more host computers.

Serial transmission

A data communications mode in which bits are sent in sequence through a single signal path.

Single-mode fiber

A step index optical fiber in which light propagates coherently, avoiding signal degradation by distance.

Snapshot

A Sun Enterprise Volume Manager feature for enabling on-line backups. Snapshots are read-only copies of the data to be backed up.

Striping

Spreading, or interleaving, logical contiguous blocks of data across multiple independent disk spindles. Striping allows multiple disk controllers to simultaneously access data, improving performance.

Switch

The name of an implementation of the fabric technology.

Throughput

A measure of sequential I/O performance, quoted as MB/second. See IOPS and I/O rate.

Topology

The components used to connect two or more ports together. Also, a specific scheme of connecting those components. Point-to-point, fabric, and arbitrated loop are example topologies.

Transceiver

A transmitter/receiver module.

Transfer rate

The rate at which bytes or bits are transferred. Usually measured in MB/second.

Volume

A volume is a virtual disk into which a file system, DBMS, or other application can place data. A volume can physically be a single disk partition or multiple disk partitions on one or more physical disk drives. Applications that use volumes do not need to be aware of their underlying physical structure. Software handles the mapping of virtual partition addresses to physical addresses.

References

Sun Microsystems Computer Company posts product information in the form of data sheets, specifications, and white papers on its Internet World Wide Web Home page at: *http://www.sun.com*.

Look for the these and other Sun technology white papers:

Evaluating Performance Through Benchmarks, White Paper, Sun Microsystems Computer Company.

Fibre Channel Technology from Sun Microsystems, Technical Brief, Sun Microsystems Computer Company, March 1997.

Sun Enterprise Volume Manager 2.5 User's Guide, Sun Microsystems Computer Company, August 1997, part number 805-1603-10.

For more information about Fibre Channel:

The Fibre Channel Association (FCA)
12407 MoPac Expressway
North 100-357
P.O. Box 9700
Austin, Texas 78766-8422
800-272-4618
http://www.fibrechannel.com

The Fibre Channel Loop Community (FCLC)
Post Office Box 2161
Saratoga, California 95070
(408) 867-1385
http://www.fcloop.org

Sun Microsystems, Incorporated
901 San Antonio Road
Palo Alto, CA 94303 USA
650 960-1300
FAX 650 969-9131
http://www.sun.com

Sales Offices

Argentina: +54-1-317-5600
Australia: +61-2-9844-5000
Austria: +43-1-60563-0
Belgium: +32-2-716-7911
Brazil: +55-11-5181-8988
Canada: +905-477-6745
Chile: +56-2-638-6364
Colombia: +571-622-1717
Commonwealth of Independent States:
+7-502-935-8411
Czech/Slovak Republics:
+42-2-205-102-33
Denmark: +45-44-89-49-89
Estonia: +372-6-308-900
Finland: +358-9-525-561
France: +33-01-30-67-50-00
Germany: +49-89-46008-0
Greece: +30-1-680-6676
Hong Kong: +852-2802-4188
Hungary: +36-1-202-4415
Iceland: +354-563-3010
India: +91-80-559-9595
Ireland: +353-1-8055-666
Israel: +972-9-956-9250
Italy: +39-39-60551
Japan: +81-3-5717-5000
Korea: +822-3469-0114
Latin America/Caribbean:
+1-650-688-9464
Latvia: +371-755-11-33
Lithuania: +370-729-8468
Luxembourg: +352-491-1331
Malaysia: +603-264-9988
Mexico: +52-5-258-6100
Netherlands: +31-33-450-1234
New Zealand: +64-4-499-2344
Norway: +47-2218-5800
People's Republic of China:
Beijing: +86-10-6849-2828
Chengdu: +86-28-678-0121
Guangzhou: +86-20-8777-9913
Shanghai: +86-21-6247-4068
Poland: +48-22-658-4535
Portugal: +351-1-412-7710
Russia: +7-502-935-8411
Singapore: +65-438-1888
South Africa: +2711-805-4305
Spain: +34-1-596-9900
Sweden: +46-8-623-90-00
Switzerland: +41-1-825-7111
Taiwan: +886-2-514-0567
Thailand: +662-636-1555
Turkey: +90-212-236 3300
United Arab Emirates:
+971-4-366-333
United Kingdom: +44-1-276-20444
United States: +1-800-821-4643
Venezuela: +58-2-286-1044
Worldwide Headquarters:
+1-650-960-1300